# Enhancing Transferability of Deep Reinforcement Learning-Based Variable Speed Limit Control Using Transfer Learning

Zemian Ke, Zhibin Li, Zehong Cao, *Member, IEEE*, and Pan Liu

***Abstract*—The study aims to evaluate the performance of the transfer learning algorithm to enhance the transferability of a deep reinforcement learning-based variable speed limits (VSL) control. The Double Deep Q Network (DDQN)-based VSL control strategy is proposed for reducing total time spent (TTS) on freeways. A real merging bottleneck is developed in the simulation and considered for the VSL control as the source scenario. Three types of target scenarios are considered, including the overspeed scenarios, adverse weather scenarios, and diverse capacity drop scenarios. A stable testing demand and a fluctuating testing demand are adopted to evaluate the effects of VSL control. The results show that by updating the neural networks, the transfer learning in the DDQN-based VSL control agent successfully transfers knowledge learned in the source scenario to other target scenarios. With the transfer learning, the entire training process is shortened by 32.3% to 69.8%, while keeping a similar maximum reward level, as compared to the VSL control with full learning from scratch. With the transferred DDQN-based VSL strategy, the TTS is reduced by 26.02% to 67.37% with the stable testing demand and 21.31% to 69.98% with the fluctuating testing demand in various scenarios, respectively. The results also show that when the task similarity between the source scenario and target scenario is relatively low, the transfer learning could lead to local optimum and may not achieve the global optimal control effects.**

***Index Terms*—Bottleneck, congestion, reinforcement learning, travel time, transferability.**

## I. Introduction

AT RECURRENT bottlenecks on the freeway, traffic demand exceeding roadway capacity could cause congestion and trigger capacity drop [1]. In turn, the capacity drop would exacerbate congestion and result in increased system travel time [2]. Variable speed limit (VSL), proposed originally to harmonize traffic flow for safety concerns, has emerged as a mainstream traffic flow control (MTFC) method to mitigate congestion through avoiding capacity drop [3]–[6]. The upstream speed limit posted on VSL sign changes according to real-time traffic conditions to adjust the flow, and thereby capacity drop is prevented at the downstream bottleneck, and congestion can be alleviated [7]–[16]. The performance of VSL control depends on the timely adjustment of the speed limit. In various environments, traffic dynamics can be greatly different so that the optimal VSL control policy could change. Hence, it is desirable to develop VSL strategies that can quickly adapt to new environments and have good transferability in different scenarios.

The most commonly used VSL control strategies are the optimal control approaches [5]–[9] and the feedback-based approaches [10]–[16]. The optimal control approaches treat the VSL control as a constrained discrete-time optimal problem. The performance relies on the accuracy of traffic flow models that represent traffic dynamics for a specific scenario [17]. When transferred to a new scenario, the approaches need to recalibrate a series of traffic models to maintain good performance, which requires large traffic flow data and burdensome model development workload. In the feedback-based VSL strategies, the feedback controllers adjust speed limits to keep the bottleneck density around the setpoint. The performance is significantly affected by the setpoint and controllers' gains, which determine the response speed [10], [11]. When transferred to a new scenario, the setpoint and controller gains should be updated for new traffic dynamics, which requires specific traffic flow analysis and complex parameter tuning.

In recent years, the reinforcement learning (RL) approaches have been applied in the freeway traffic control tasks, and have attracted significant attention because of its good performance [18]–[22]. In a RL-based VSL, the agent perceives the traffic dynamics through interacting with the traffic environment, and is able to develop an optimal control policy through adequate training from the state-action-reward outcomes [18], [19]. The optimal policy obtained from the original scenario, if applied to a new scenario directly, may lead to decreased control performance as the state-action-reward pairs in the new scenario are different from those in the original scenario [19]. The common way for a new scenario is to create a new RL-based VSL with full learning (learning

Manuscript received October 25, 2019; revised February 11, 2020; accepted April 9, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 71871057, in part by the Fundamental Research Funds for the Central Universities under Grant 2242019R40060, and in part by Southeast University, through the Zhongying Young Scholars Project. The Associate Editor for this article was A. Jolfaei. *(Corresponding authors: Zhibin Li; Zehong Cao.)*

Zemian Ke, Zhibin Li, and Pan Liu are with the School of Transportation, Southeast University, Nanjing 210096, China (e-mail: kezemian@foxmail.com; lizhibin@seu.edu.cn; liupan@seu.edu.cn).

Zehong Cao is with the Discipline of Information and Communication Technology (ICT), University of Tasmania, Hobart, TAS 7005, Australia (e-mail: zehong.cao@utas.edu.au).

Digital Object Identifier 10.1109/TITS.2020.2990598

from scratch), which requires a large computing workload and long training process.

Transfer learning is concerned with the connection between training in different but related scenarios [23]–[27]. It could shorten the training process in one scenario by utilizing existing knowledge learned from another related scenario [23]. Although traffic parameters in various scenarios are different, there are still some attributes in common such as the correlation between traffic flow variables [28] and the impacts of control actions on traffic operation [5]. As a result, through incorporating the transfer learning with the RL-based control strategy, it is believed that the transferability of the control strategy can be enhanced in terms of quick adaption to new environments and maintenance of good performance. The study of Kreidieh *et al.* (2018) confirmed the benefits of transferring the polices developed from a closed network to open network tasks [27]. However, to the best of our knowledge, until now, no studies have evaluated the practicability of incorporating the transfer learning with the VSL control tasks.

This study aimed to integrate the transfer learning algorithm with the deep reinforcement learning-based VSL control strategy for improving its transferability. We first trained the Double Deep Q Network (DDQN)-based VSL for a source scenario. Then we developed three types of target scenarios, including the overspeed scenarios, the adverse weather scenarios, and the diverse capacity drop scenarios, for testing the transferability of the DDQN-based VSL strategy. The rest of the paper is organized as follows. Section 2 illustrates the DDQN-based VSL and the incorporation of transfer learning. Section 3 shows the development of the simulation model. Section 4 introduces the experimental setup of various scenarios. The performances of the DDQN-based VSL control are analyzed in section 5. Finally, the main conclusions are drawn in section 6.

## II. Methodology

### A. DDQN-Based VSL Control Strategy

*1) Statement of VSL Control Problem:* A typical VSL control scheme at a merging bottleneck is illustrated in Fig. 1 (a). The VSL system includes two sections: (1) An upstream VSL controlled section, in which the outflow is controlled by adjusting the posted speed limits; and (2) an acceleration section, which allows vehicles to accelerate from low speed within the controlled section to roughly critical speed as they reach the bottleneck [5], [10]. The VSL control could adjust the mainline traffic flow entering downstream to mitigate or eliminate the capacity drop at the bottleneck. Thus, the travel time can be reduced by improving the outflow rate. The key task of a VSL control strategy is how to determine the optimal speed limit for a given traffic state in the freeway area.

*2) DDQN-Based VSL Control Strategy:* A DDQN-based VSL control strategy includes three modules (see Fig. 2): VSL simulation, DDQN agent, and neural networks [29]. The optimal speed limit for a particular traffic state is determined by the DDQN agent. The agent perceives the traffic state $s_t$ and selects a speed limit $a_t$. The speed limit posted in the controlled section leads to the transition of traffic state on the

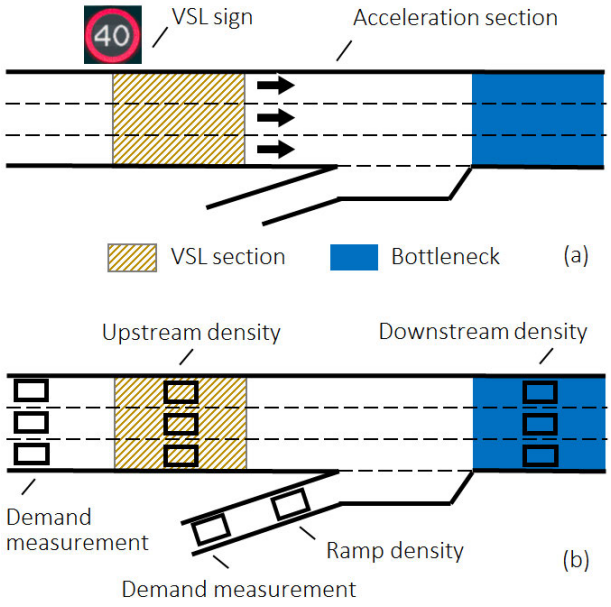


Fig. 1. (a) A freeway merging bottleneck with the VSL control; (b) Density and demand measurements in the DDQN-based VSL control.

freeway. The agent receives a reward $R_{t+1}$ for the new state $s_{t+1}$, and then the transition ($s_t$, $a_t$, $R_{t+1}$, $s_{t+1}$) is stored into memory to update parameters $\theta$ of the neural network using the Bellman equation. In the DDQN algorithm, the neural network works as a function approximator to estimate optimal Q (action) values $Q*(s, a)$. For the neural network with parameters $\theta_i$, as updating step $i \to \infty$, the neural network estimation $Q(s, a; \theta_i) \to Q*(s, a)$. Then the optimal speed limit $a*$ can be obtained as $a* \equiv \text{argmax}_a Q*(s, a)$.

There are five crucial elements in the DDQN-based VSL control strategy as follows.

**a) Action**. The VSL controls mainstream traffic flow by adjusting the speed limit, which is considered as the action in the DDQN-based VSL control. In the real-world cases, the variable message sign can only post discrete speed limits with an increment of 5 mph. The maximum speed limit is 65mph for the study freeway which would be illustrated in the following sections. Therefore, the action set is given by {5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65} mph.

**b) State**. In the DDQN algorithm, traffic state space is represented by several continuous variables that should be able to reflect the traffic state of the freeway. In the present study, to depict the traffic flow conditions under the influence of VSL control, five continuous variables are adopted: demand flow of upstream mainline; demand flow of on-ramp; density at the downstream bottleneck (i.e., the immediate downstream of the merging area); density at the upstream VSL area; and density on the on-ramp (see Fig. 1 (b)). In addition, to help the agent perceive the speed limit change rate along time, the speed limit at the former time step is also included in the state. Finally, a six-dimensional vector is used to represent the state.

**c) Reward**. The objective of VSL control is to reduce the system travel time. Consider a freeway system that consists of several origins $I$ and destinations $I'$, and a discrete-time representation of traffic variables with time index $k$ and time

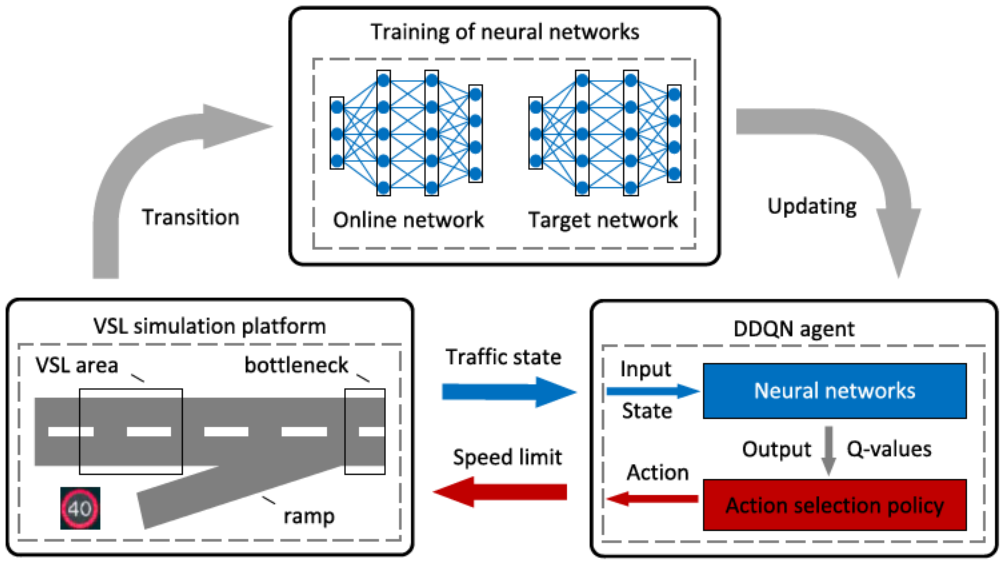


Fig. 2. Flowchart of the DDQN-based VSL control strategy.

interval $\eta$. According to [30], the total time spent (TTS) over a time horizon $K$ can be calculated by:

$$TTS = \eta \sum_{k=1}^{K} N(k) \tag{1}$$

where $N(k)$ is the total number of vehicles in the network at time $k$. Due to the conservation of vehicles:

$$TTS = \eta \sum_{k=1}^{K} \left[ N(0) + \eta \sum_{\kappa=0}^{k-1} q(\kappa) - \eta \sum_{\kappa=0}^{k-1} s(\kappa) \right] \tag{2}$$

where $N(0)$ is the total number of vehicles in the network at the beginning, $q(\kappa)$ is the total arriving flow from origins at time $\kappa$, and $s(\kappa)$ is the total exit flow from destinations at time $\kappa$. Assuming that the arriving flow $q(\kappa)$ and its spatial and temporal distribution are independent of any control, the minimization of travel time in the system can be achieved by maximizing the exit flow [1], [2], [28], [30]–[32]. For the merging bottleneck in Fig. 1 (a), the reduction in TTS is mainly determined by the bottleneck discharge flow with the VSL. Traffic flow can be represented by density, and each density corresponds to a unique flow [28], [30], [31], [33], [34]. Density (or occupancy) has been considered as an indicator for traffic state [35], [36], and has been the control objective in many ramp metering algorithms [30], [31] and VSL algorithms [10], [11], [37]. At the bottleneck area, maximum exit flow is reached when density is equal to its critical value. Thus, in our study, the reward function is determined according to the density measured at the immediate downstream of the bottleneck. The largest reward is received if the downstream density equals critical value. The reward decreases linearly as the downstream density deviates from the critical value (see Fig. 3). The reward $r$ is calculated as follows:

$$r = \begin{cases} c \cdot d, & if\ d < d_c \\ c \cdot d_c - c \cdot (d - d_c), & if\ d \geq d_c \end{cases} \tag{3}$$

where $d$ denotes the downstream density (veh/mi/ln), $d_c$ denotes the critical density (which is 26.75 veh/mi/ln in the present study), and the parameter $c$ is set as 0.02. To stimulate an agent to achieve the critical density and increase the algorithm convergence speed [19], an extra incentive of 0.5 is added to the reward of states near to critical density, if $26 \leq d \leq 27.5$. A penalty of 0.5 is added to the reward for severely congested states, if $d \geq 45$. When the upstream mainline density and the downstream mainline density are both smaller than 25 veh/mi/ln, indicating a free flow condition, a penalty of 0.2 is added to the reward if the maximum speed limit has not been chosen. For the sake of safety, the difference between speed limits at consecutive time steps cannot be too large. Therefore, if the last speed limit minus the chosen speed limit is greater than 10 mph, a penalty of 0.1 is added to the reward. The aforementioned incentive and penalty values are carefully decided from multiple rounds of simulation tests. The discount factor $\gamma$ decides the agent's ability to account for future rewards, and a value of 0.8 is adopted [19].

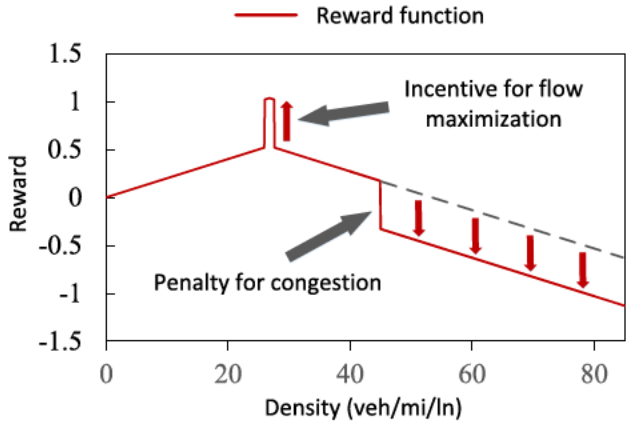


Fig. 3. Reward function of the DDQN-based VSL strategy.

**d)Neural Networks**. Mnih *et al.* (2015) found the use of the target network would increase the stability and performances of the algorithm dramatically [38], so two neural networks (the online network and the target network) are utilized to

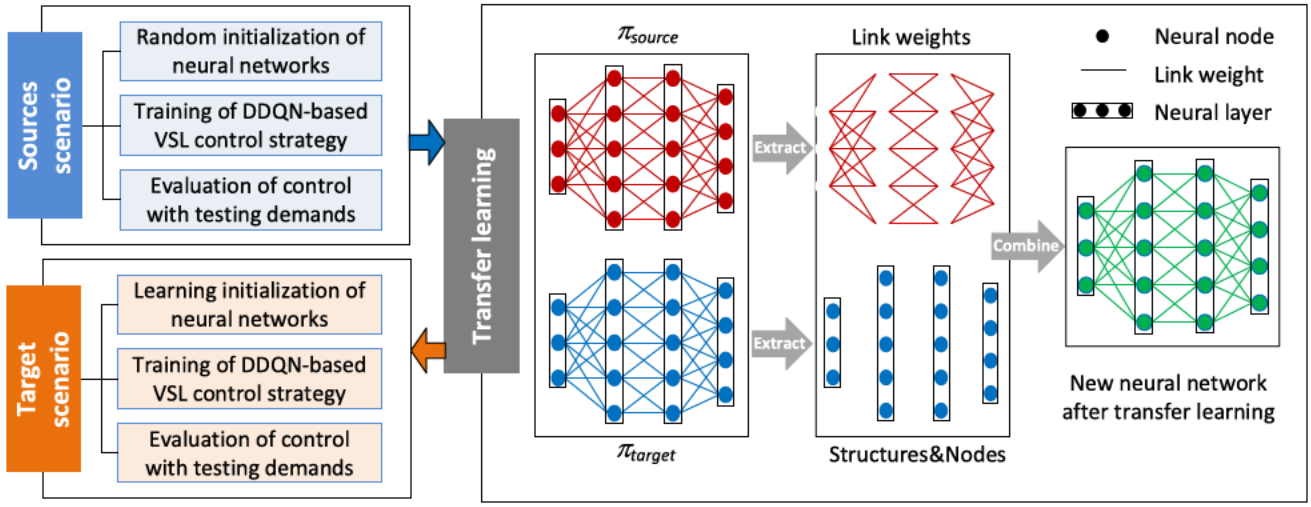


Fig. 4. Flowchart of the transfer learning in the VSL control.

estimate Q-values of different speed limits for a given traffic state. The target network is a historical version of the online network. The input of both neural networks is the six-dimensional traffic state vector, and the output is Q-values of different speed limits. After sufficient training, for any given traffic state, the neural networks are able to estimate Q-values of speed limits even though some traffic states may never be encountered. With the well-trained neural networks, the DDQN agent could realize the optimal policy by choosing the optimal speed limit with the maximal Q-value for an given state.

**e)Action selection policy**. The action selection policy applies the $\varepsilon$-greedy method, which is able to make a balance between exploitation and exploration. In the method, the agent could choose an action with the largest Q-value most of the time (exploitation), or select an action randomly from all the actions (exploration). In our study, the $\varepsilon$ is set to be large at the beginning of training to explore more potential. As the training processes, the $\varepsilon$ anneals to a small value to exploit more from the existing learned knowledge. During the testing (implementation) process, the difference between two consecutive speed limits posted on VSL is limited to 10 mph to avoid abrupt speed limit change for safety concerns.

### *B. Integration of Transfer Learning With VSL*

Transfer learning is able to utilize the knowledge acquired from training in one scenario to improve the training in other related scenarios so that the training process can be shortened significantly [23]–[27]. In general, the original scenario is treated as the source scenario, and closely related new scenarios can be treated as the target scenarios to conduct the transfer learning.

In the source scenario, the neural networks of DDQN are initialized randomly. The neural networks in the target scenario ($\pi_{target}$) are initialized based on the well-trained network from the source scenario ($\pi_{source}$). Two mappings from the target scenario to the source scenario are provided: $\chi_{H,X}$ maps each state variable in the target scenario to the most similar state variable in the source scenario: $\chi_{H,X}(x_{i,target}) = x_{j,source}$, and $\chi_{H,A}$ maps each action in the target scenario to the most similar action in the source scenario: $\chi_{H,A}(a_{i,target}) = a_{j,source}$. At first, $\pi_{target}$ is defined to have no links, one input node for each state variable, one output node for each action, and the same number of hidden nodes as in $\pi_{source}$. Thus, each node $n$ in $\pi_{target}$ can be mapped back to a node in $\pi_{source}$ via a function $\psi$:

$$\psi(n) = \begin{cases} \chi_{H,X}(n), & \text{if } n \text{ is an input node} \\ \chi_{H,A}(n), & \text{if } n \text{ is an output node} \\ n, & \text{if } n \text{ is a hidden node} \end{cases} \tag{4}$$

Accordingly, $\pi_{target}$ can be generated through copying the links that connect the corresponding nodes in $\pi_{source}$. For each pair of nodes $n_i$, $n_j$ in $\pi_{target}$, if a link exists between $\psi(n_i)$ and $\psi(n_j)$ in $\pi_{source}$, a new link with the same weight is created between $n_i$ and $n_j$. By applying this method, $\pi_{target}$ can be initialized with prior knowledge from the source scenario.

The procedure of incorporating the transfer learning with the DDQN-based VSL strategy is shown in Fig. 4. Three components are contained: (1) A DDQN-based VSL strategy designed for the source scenario is initialized randomly, and then is trained in the source scenario; (2) With the transfer learning enabled, the trained DDQN-based VSL is transferred to adapt to the target scenario; (3) The DDQN-based VSL directly acquired in part (2) is initially adopted, and then is trained in the target scenario. The performance of DDQN-based VSL in the target scenario is evaluated using testing demands.

## III. Simulation Network

In this study, we develop a macroscopic simulation platform based on the Cell Transmission Model (CTM). The focus of our study is on the macroscopic traffic variables (such as density, flow, and average speed) and state dynamics at the merging bottleneck. We do not look into the details of the microscopic car-following or lane-changing behaviors of

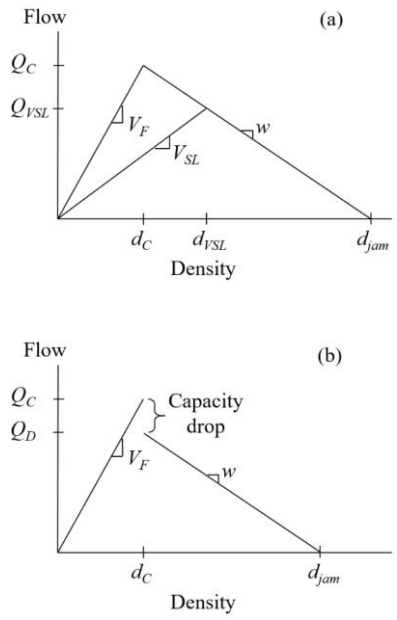


Fig. 5. (a) Fundamental diagram in CTM; (b) Fundamental diagram with the capacity drop.

individual vehicles. In addition, the decision of VSL control is made on the basis of aggregated traffic data collected from loop detectors. Detailed information about individual vehicles is actually not used. Furthermore, our experiment involves a large number of iterative calculations. The computing workload and time of the microscopic simulation model are much larger and longer than that of the CTM. As a result, instead of using microscopic simulation models, the CTM is considered fitting the purpose of our study better.

By dividing the corridor into sub-sections (i.e., cells), CTM predicts the macroscopic traffic characteristics by evaluating the flow and density at a finite number of intermediate points at different time steps [33]. The traffic in each cell operates according to the fundamental diagram, which is approximated by a triangular shape. To build the simulation platform for VSL control, several modifications are made to the traditional CTM [19], [39], [40].

The adopted fundamental diagram is shown in Fig. 5 (a). The left limb of the triangle represents the sending function, and the right limb represents the receiving function. For cell $i$, the sending function represents the vehicles that can supply to the downstream cell $i+1$ with a flow rate of $\sigma_i(k)$, where $k$ is the time step. The receiving function represents the available space in cell $i$ which determines how many vehicles can enter cell $i$ from the upstream cell $i-1$ with a flow rate of $\delta_i(k)$. With the VSL control, the sending and receiving functions are determined by the minimum value between the speed limit $V_{SL}$ and the free flow speed $V_F$:

$$\sigma_i(k) = \min\{\min\{V_{SL}(k), V_F\} \cdot d_i(k) \cdot n_i, Q_{VSL}\} \quad (5)$$

$$\delta_i(k) = \min\{w_i \cdot (d_{i,jam} - d_i(k)) \cdot n_i, Q_{VSL}\} \quad (6)$$

where $d_i(k)$ is the density at cell $i$ at time $k$, $n_i$ is the number of lanes, $Q_{VSL}$ is the maximum flow under the current speed limit, $w_i$ is the kinematic wave speed, and $d_{i,jam}$ is the jam density. The evolution of density and speed within each cell are calculated according to the flow rate between cells, and the flow rate is determined as the minimum value of the sending and receiving functions.

The discharge flow drops below the bottleneck capacity after congestion forms [1], [2]. To model the capacity drop, it is assumed that the bottleneck cell is characterized by an inverse $\lambda$-shaped fundamental diagram (see Fig. 5 (b)). The flow is calculated by the left limb before capacity drop occurs and is calculated by the right limb after capacity drop occurred. Note that because capacity drop does not affect free flow speed, the size of the bottleneck cell is not influenced by capacity drop, and remains constant during simulation. The sending function with the capacity drop is determined by

$$\sigma_i(k) = \begin{cases} V_F \cdot d_i(k) \cdot n_i, & if\ d_i(k) \le d_C \\ Q_D, & if\ d_i(k) > d_C \end{cases} \quad (7)$$

where $Q_C = V_F \cdot d_C \cdot n_i$ is the capacity of the bottleneck (veh/h) before the capacity drop, $Q_D$ is the maximum discharge flow rate (veh/h) after capacity drop, and $d_C$ is the critical density.

Four traffic flow parameters are considered for the calibration of the fundamental diagram, including the free flow speed, the capacity flow, the discharge flow after capacity drop, and the speed of kinematic wave [33]. The capacity flow and discharge flow after capacity drop are calculated using the cumulative vehicle count curves at the bottleneck location [41]. The speed of the kinematic wave is calculated by monitoring the traffic states at the detector stations located upstream of the active bottleneck [42].

## IV. Experiment Design

### A. Setup of Source Scenario

The source scenario is a 6-mile long freeway section of the northbound of Interstate 880 freeway in Oakland, United States. A merging recurrent bottleneck locates at the downstream of this freeway section (see Fig. 6), and is activated during both morning and afternoon peak periods. We expected to enhance the throughput of this recurrent bottleneck by implementing VSL control at the upstream.

The freeway section is simulated by the modified CTM illustrated before. One-month traffic data are obtained from the Highway Performance Measurement System (PeMS) [43] for calibrating the parameters in the CTM and validation. For details of the calibration and validation, we refer the readers to [19], [40]. The calibration results suggest the free flow speed is 65 mph, the capacity of the freeway mainline before the capacity drop is 1750 veh/h/ln, the magnitude of the capacity drop is 8.4%, and the speed of the kinematic wave is 9.5 mph. The simulation results are compared with the field data for validation. The Mean Absolute Percent Error (MAPE) of simulated flows is 9.2%, and the MAPE of simulated speeds is 11.2%.

To implement VSL control in the CTM, the VSL area and control period should be defined. The location and length of VSL area were determined based on acceleration distance,

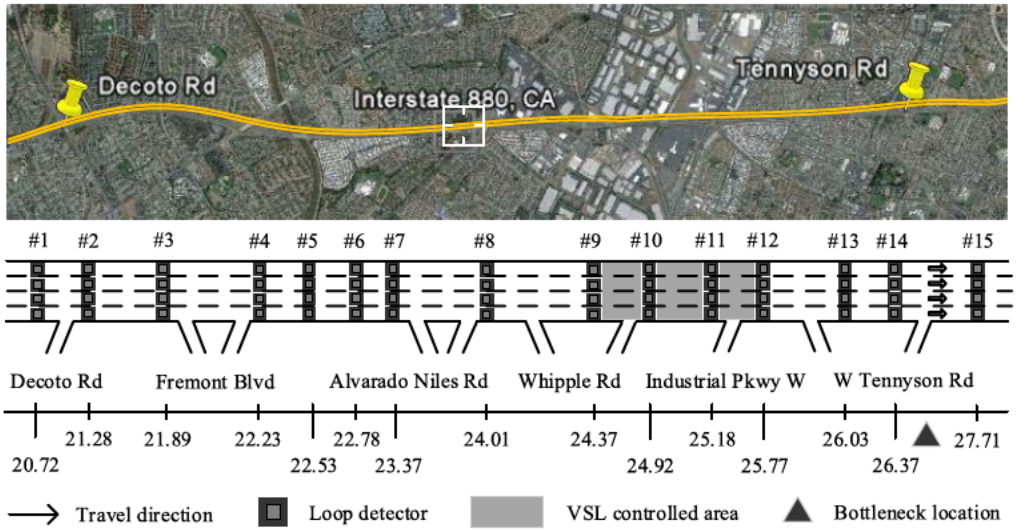


Fig. 6. Illustration of the study freeway section.

TABLE I
SCENARIO DESIGN IN PRESENT STUDY

| Scenario | $V_O$ [a] (mph) | $V_F$ [b] (mph) | $C_{drop}$ [c] |
|---|---|---|---|
| Source scenario | 0 | 65 | 8.1% |
| *Driver overspeed scenarios* | | | |
| Target scenario 1 | 5 | 65 | 8.1% |
| Target scenario 2 | 7.5 | 65 | 8.1% |
| Target scenario 3 | 10 | 65 | 8.1% |
| *Adverse weather scenarios* | | | |
| Target scenario 4 | 0 | 60 | 8.1% |
| Target scenario 5 | 0 | 57.5 | 8.1% |
| Target scenario 6 | 0 | 55 | 8.1% |
| *Diverse capacity drop scenarios* | | | |
| Target scenario 7 | 0 | 65 | 5% |
| Target scenario 8 | 0 | 65 | 15% |

[a] $V_O$ denotes magnitude of overspeed.
[b] $V_F$ denotes free flow speed.
[c] $C_{drop}$ denotes magnitude of capacity drop.

deceleration distance, and field detector distribution. To maximize outflow at the bottleneck, speed around bottleneck should be kept at the critical speed. However, the speed at the VSL area may be lower than the critical speed, so an acceleration area is needed to allow vehicles to accelerate from the lowest speed in the VSL area to critical speed at bottleneck [5]. Besides, the VSL zone length should allow vehicles to decelerate from upstream free flow speed to the posted lowest speed limit in the zone. In addition, the beginning and end of the VSL zone should have detectors so that the traffic state of the VSL area can be monitored. Thus, in our study freeway, after satisfying the aforementioned requirements, the VSL area was determined as the gray area in Fig. 6, so the available acceleration distance was 0.6 miles, and the VSL zone length was 1.4 miles. As for the control period, too long control period would limit control effects, and too short control period may make drivers confused. Therefore, we tested several control periods and found the control period of 30 s lead to the best control effects. Finally, the control period was set to be 30 s.

As mentioned before, the traffic state of the DDQN-based VSL consisted of the speed limit of the last time step and five traffic variables that could be collected from freeway detectors. For the study freeway, the mainline demand flow was obtained from detector #1 (see Fig. 6). The density of the upstream VSL area was the average of densities calculated by measured data of detectors #9-#12. The downstream bottleneck density was the density of detector #15. The ramp demand was the flow into the on-ramp of W Tennyson Rd, and the ramp density was the density from the detector located at the ramp.

### B. Design of Target Scenarios

The transferability of the DDQN-based VSL is evaluated by transferring the control strategy from the source scenario to several target scenarios. In the present study, three types of target scenarios are developed, including the overspeed scenarios, the adverse weather scenarios, and the diverse capacity drop scenarios. All target scenarios are, to some extent, related to the source scenario, and are developed based on observations in [2], [44] and [45]. The experiment design for various scenarios is shown in Table I.

## V. RESULTS OF SIMULATION ANALYSIS

### A. Training in Source Scenario

The DDQN-based VSL was trained through interacting with the CTM, by which neural networks were updated. At the $t$-th time step, the DDQN-based VSL received traffic state $s_t$ from the CTM, and then chose a speed limit $a_t$ to be posted in the CTM. The new simulating traffic state $s_{t+1}$ after the speed limit had been implemented and the reward $R_{t+1}$ for the state transition was transmitted to the DDQN-based VSL, and then the transition ($s_t$, $a_t$, $R_{t+1}$, $s_{t+1}$) was saved in memory for updating of neural networks.

To improve the training effect, the training process was broken down into a sequence of separate episodes [46]. Each episode included 531 time steps of 30s. In addition, to prevent the overfitting of neural networks brought by overtraining,

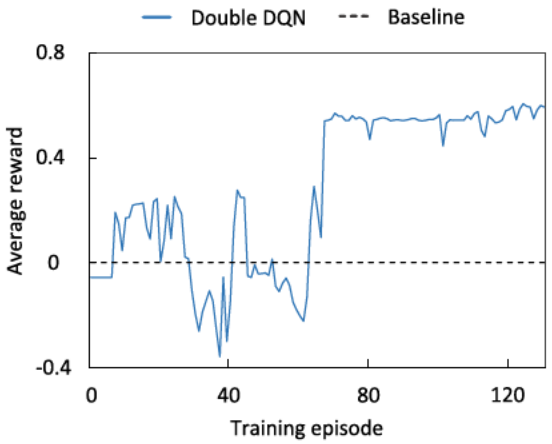


Fig. 7. Reward acquired by the DDQN-based VSL and baseline in the source scenario.

an appropriate stopping criterion was required to stop training opportunely. A general stopping criterion is when the gain of reward is positive and becomes sufficiently small [46]. Average reward ($r_n$) received at the $n$-th episode was recorded, and the gain ($g_n$) of $r_n$ was calculated as:

$$g_n = \left(\sum_{i=n-k+1}^{n} r_i - \sum_{i=n-2k+1}^{n-k} r_i\right) \Big/ \sum_{i=n-2k+1}^{n-k} r_i \tag{8}$$

Thus, the training process would stop once $g_n$ was positive and sufficiently small. After multiple tests, the value of $k$ was set to be 10, and the training was stopped when $g_n$ was positive and less than 5%.

The DDQN-based VSL was trained, and the reward curve is shown in Fig. 7. After a fluctuation at the first 60 episodes, $r_n$ increased significantly and reached a peak after around 120 episodes. The stopping criterion was satisfied at the 131$^{th}$ episode, and training was ended. The reward for the baseline situation when no VSL control was applied is also shown in Fig. 7. It is identified that after adequate training, the reward of DDQN-based VSL is much higher than the baseline, suggesting that the DDQN agent has learned the optimal policy for reward maximization through appropriate training.

The trained DDQN-based VSL was tested on the CTM using two testing demands, including a stable demand and a fluctuating demand (see Fig. 8). The TTS in the stable and fluctuating demand was denoted as $TTS_s$ and $TTS_f$, respectively. The situation without any control was also considered for comparison purposes. The simulation results show that in the no control case, $TTS_s$ and $TTS_f$ are 1020 veh · h and 719 veh · h respectively. When the DDQN-based VSL is applied, $TTS_s$ and $TTS_f$ decrease to 812 veh · h and 587 veh · h, indicating a 50.47% reduction in $TTS_s$ and a 35.53% reduction in $TTS_f$. The results suggest that the trained DDQN-based VSL is able to reduce travel time effectively in the source scenario.

### *B. Testing of Transfer Learning in Overspeed Scenarios*

Drivers comply strictly with the speed limit in the source scenario, which may not be realistic in actual situations. The driver compliance to speed limit was reported to have an obvious impact on the safety and operational effects of VSL control [44]. In our simulation model, a new variable called driver overspeed magnitude is introduced [19], which is given by:

$$V'_{SL} = \min(V_F, V_{SL} + V_O) \tag{9}$$

where $V'_{SL}$ is the actual traffic speed when the speed limit is $V_{SL}$, $V_F$ is the free flow speed, and $V_O$ is the magnitude of overspeed. Speed data collected by [44] showed that only 8% vehicles are speeding with overspeed larger than 12 mph. Therefore, the magnitude of overspeed was set to be 5 mph, 7.5 mph or 10 mph (scenario 1-3 in Table I).

The training processes of the DDQN-based VSL with transfer learning (VSL-TL) in three overspeed scenarios were shown in Fig. 9. For comparison purposes, we also showed a DDQN-based VSL with full learning (VSL-FL), i.e., training from scratch. The results are shown in Fig. 9. It is clearly identified that in the three scenarios, both VSL strategies were able to reach convergence (i.e., gain stable and maximum reward) after adequate training. However, transfer learning greatly fastened the training process. More specifically, in scenario 1, the VSL-FL received an unstable and low reward in the initial 80 episodes. The reward increased remarkably at the 95th episode, and became stable after the 120th episode. The stopping criterion was reached at the 141th episode. While for the VSL-TL, reward increased rapidly after the 15th episode and became stable after the 40th episode. The stopping criterion was satisfied at the 61th episode. Similar trends were also found in scenario 2 and 3. The training time has been shortened by 56.7%, 44.4%, and 32.3% by the transfer learning in the three scenarios. Note that in the scenario with a larger magnitude of overspeed, the transfer learning showed less reduction in the training time, which is reasonable because as the difference between the source and the target scenario increases, the knowledge learned from the source scenario contribute less to the training process in the target scenario [22].

After training, the control effects of the VSL-TL and the VSL-FL were evaluated in two testing demands (see Fig. 8), and the results are shown in Table II. The control effects of the VSL-TL and the VSL-FL were compared with the no control situation. With the VSL-TL, the $TTS_s$ was reduced by 42.75% to 50.78%, and the $TTS_f$ was reduced by 31.03% to 35.77%. With the VSL-FL, the $TTS_s$ was reduced by 42.73% to 50.67%, and the $TTS_f$ was reduced by 31.11% to

TABLE II
EFFECTS OF THE VSL CONTROL IN THE OVERSPEED SCENARIOS

| Control strategy | $TTS_s$ (veh·h) | $TTS_s$ reduction | $TTS_f$ (veh·h) | $TTS_f$ reduction |
|---|---|---|---|---|
| No control | 1020.68 | - | 719.34 | - |
| *Scenario 1 ($V_O$ = 5 mph)* | | | | |
| VSL-TL | 502.39 | 50.78% | 462.04 | 35.77% |
| VSL-FL | 503.48 | 50.67% | 463.10 | 35.62% |
| *Scenario 2 ($V_O$ = 7.5 mph)* | | | | |
| VSL-TL | 584.34 | 42.75% | 496.10 | 31.03% |
| VSL-FL | 584.56 | 42.73% | 495.54 | 31.11% |
| *Scenario 3 ($V_O$ = 10 mph)* | | | | |
| VSL-TL | 553.41 | 45.78% | 462.25 | 35.74% |
| VSL-FL | 553.02 | 45.82% | 462.33 | 35.73% |

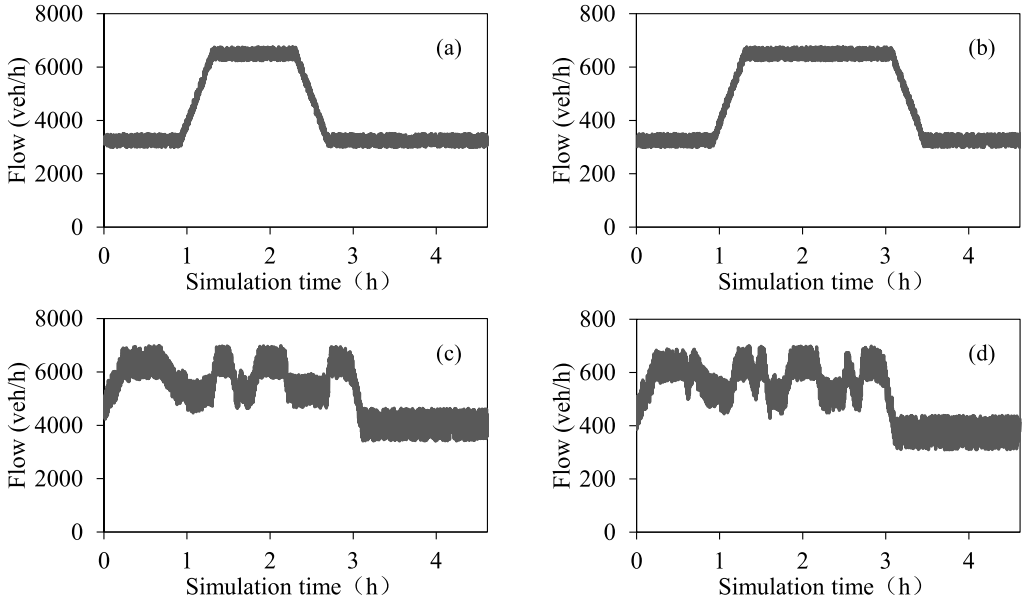


Fig. 8. Traffic operation at the bottleneck in the stable demand scenario.

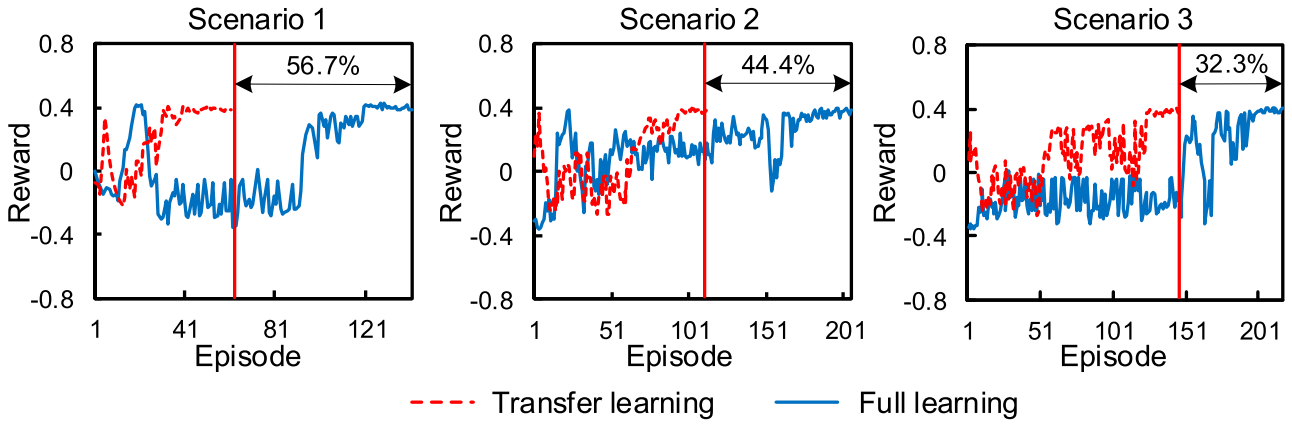


Fig. 9. Average reward acquired by the DDQN-based VSL in overspeed scenarios.

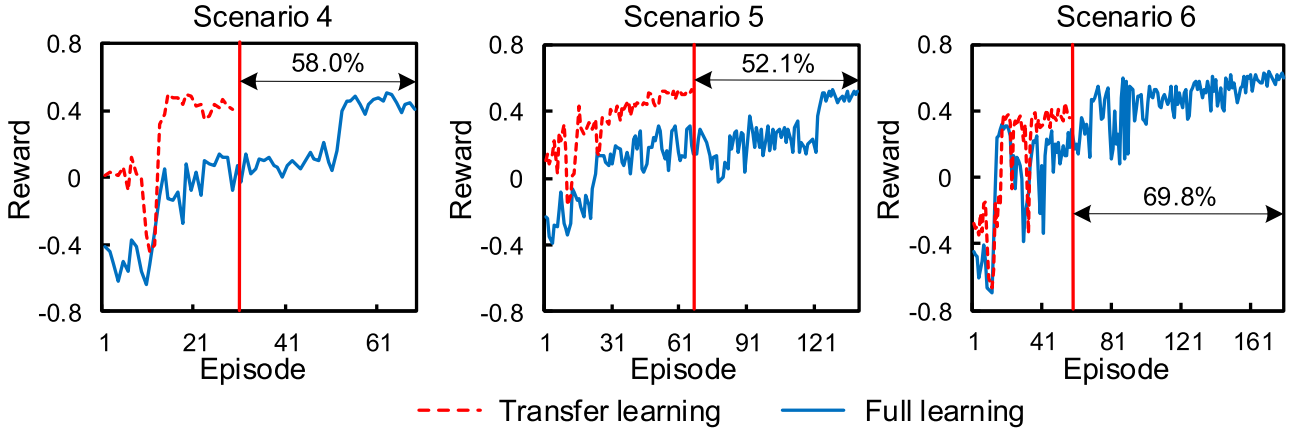


Fig. 10. Average reward acquired by the DDQN-based VSL in adverse weather scenarios.

35.73%. Therefore, the above results suggested when compared with the VSL-FL, the VSL-TL not only required a shorter training process but also realized equivalent control performances.

### *C. Testing of Transfer Learning in Adverse Weather Scenarios*

The DDQN-based VSL strategy in the source scenario was transferred to the adverse weather scenarios. The weather

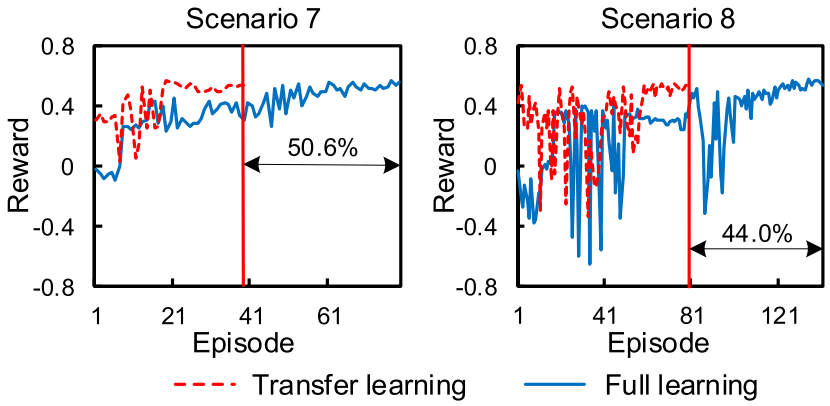


Fig. 11. Average reward acquired by the DDQN-based VSL in diverse capacity drop scenarios.

TABLE III
EFFECTS OF THE VSL CONTROL IN THE ADVERSE WEATHER SCENARIOS

| Control strategy | $TTS_S$ (veh·h) | $TTS_S$ reduction | $TTS_f$ (veh·h) | $TTS_f$ reduction |
|---|---|---|---|---|
| *Scenario 4 ($V_F$ = 60 mph)* | | | | |
| No control | 1799.28 | - | 3409.86 | - |
| VSL-TL | 1119.92 | 37.76% | 1511.23 | 55.68% |
| VSL-FL | 1121.10 | 37.69% | 1503.47 | 55.91% |
| *Scenario 5 ($V_F$ = 57.5 mph)* | | | | |
| No control | 2318.39 | - | 5500.30 | - |
| VSL-TL | 1417.63 | 38.85% | 2148.55 | 60.94% |
| VSL-FL | 1433.58 | 38.16% | 2135.79 | 61.17% |
| *Scenario 6 ($V_F$ = 55 mph)* | | | | |
| No control | 2992.45 | - | 7733.24 | - |
| VSL-TL | 2213.84 | 26.02% | 4375.51 | 43.42% |
| VSL-FL | 1876.45 | 37.29% | 3564.72 | 53.90% |

TABLE IV
EFFECTS OF THE VSL CONTROL IN THE DIVERSE CAPACITY DROP SCENARIOS

| Control strategy | $TTS_S$ (veh·h) | $TTS_S$ reduction | $TTS_f$ (veh·h) | $TTS_f$ reduction |
|---|---|---|---|---|
| *Scenario 7 ($C_{drop}$ = 5%)* | | | | |
| No control | 812.26 | - | 587.89 | - |
| VSL-TL | 502.39 | 38.15% | 462.64 | 21.31% |
| VSL-FL | 505.18 | 37.81% | 461.34 | 21.53% |
| *Scenario 8 ($C_{drop}$ = 15%)* | | | | |
| No control | 1549.31 | - | 1543.42 | - |
| VSL-TL | 505.55 | 67.37% | 463.37 | 69.98% |
| VSL-FL | 504.33 | 67.45% | 464.91 | 69.88% |

could influence driving behaviors, and thus result in changes in traffic operation [45], [47]. The free flow speed $V_F$ was reduced from 65 mph to 60mph, 57.5mph or 55mph in adverse weather (scenario 4-6 in Table I), resulting in decreases in the capacity. The critical density and jam density were assumed not affected by the weather conditions [48].

The rewards of the VSL-TL and the VSL-FL during the training process are shown in Fig. 10. Similarly, the transfer learning greatly fastened the training process: the training time before reaching the stop criterion was reduced by 58%, 52.1%, and 69.8%, respectively, in the three weather conditions. Note that in scenario 4 and 5, the VSL-TL was able to acquire reward as high as the fully trained strategy. However, in scenario 6, the transfer learning reached a stopping criterion in the 54$^{th}$ episode when the reward hasn't reached the maximum value. It indicated that transfer learning led to a local optimum other than the global optimum. Such an issue may be overcome by relaxing the stopping criterion for a longer training period.

After training, the VSL-TL and the VSL-FL were evaluated in the two testing demands. The results are shown in Table III. Compared with the baseline without VSL control, in scenario 4 and 5, the VSL-TL reduced the $TTS_s$ by 37.76% and 38.85%, and reduced the $TTS_f$ by 55.68% and 60.94%, which were almost the same as those of the VSL-FL. However, in scenario 6, the VSL-FL outperformed the VSL-TL. More specifically, the VSL-TL reduced the $TTS_s$ and the $TTS_f$ by 26.02% and 43.42%, while the VSL-FL reduced the $TTS_s$ and the $TTS_f$ by 37.29% and 53.90%. We also found that the VSL control was more effective in reducing travel time in the fluctuating demand. A possible reason was that the adverse weather resulted in more severe congestion and larger travel delay with the fluctuating demand. In such conditions, the VSL control could bring more benefits to traffic operation.

### *D. Testing of Transfer Learning in Diverse Capacity Drop Scenarios*

Empirical studies have shown that the magnitude of capacity drop at the merging bottleneck could vary in different periods or days [2]. The larger capacity drop indicates lower exiting flow and more severe congestion. Thus, it is valuable to test the transferability of the DDQN-based VSL control strategy for different capacity drop scenarios. In our study, the magnitude of capacity drop was set to be 5% and 15%, respectively, in scenarios 7 and 8 (see Table I). The fundamental diagram in the CTM was adjusted according to [39], [49].

The rewards of the VSL-TL and the VSL-FL are shown in Fig. 11. It can be seen that the transfer learning significantly reduced the training time before reaching the stopping criterion by 50.6% and 44.0% in scenario 7 and 8, respectively. The maximum reward achieved by the VSL-TL was very close to that achieved by the VSL-FL. The control results in Table IV showed that the VSL-TL reduced the $TTS_s$ and the $TTS_f$ by 38.15% and 21.31% in scenario 7, and by 67.37% and 69.98% in scenario 8, which were almost same as those of the VSL-FL. Note that the performance of the VSL in scenario

8 was much better than that in scenario 7. The main reason is that the congestion severity and TTS were greatly affected by the magnitude of the capacity drop [3]–[5], [9]. Aimed at preventing the occurrence of capacity drop and maintaining the bottleneck capacity, the VSL control is expected to be more effective in the scenario with a larger capacity drop.

## VI. Conclusion and Discussion

This study evaluated the incorporation of transfer learning to enhance the transferability of the DDQN-based VSL control strategy, which aims at reducing travel time at freeway bottleneck areas. The transfer learning in the DDQN-based VSL was able to transfer an optimal control policy for a source scenario to target scenarios, which improves the training processes in the target scenarios. Three types of target scenarios, including the overspeed scenarios, the adverse weather scenarios, and the diverse capacity drop scenarios, were developed in simulation to test the transferability performance of the DDQN-based VSL control.

In the overspeed scenarios, compared with the VSL-FL, the VSL-TL shortened the training process by 32.3% to 56.7%, while achieving an equivalent performance in reducing the TTS. In the adverse weather scenarios, the VSL-TL shortened the training processes significantly and achieved similar effects as the VSL-FL in scenarios 4 and 5. However, the VSL-TL did not achieve a global optimum in scenario 6 where the drop of speed is large. Such an issue may happen when the task similarity between the source scenario and the target scenario was relatively low [50]. However, to the best of our knowledge, quantitative methods measuring such task similarity are still not available. This is also our future work. For different capacity drop scenarios, the VSL-TL quickly converged to the optimum, shortening the training process by 50.6% and 44.0% and reducing the TTS by 21.31% to 69.98%. Note that the TTS reduction percentages in scenario 8 were large but still reasonable. It is because that the capacity drop magnitude was large, and both the VSL-TL and the VSL-FL could eliminate the capacity drop. Moreover, the total TTS was calculated in a short freeway segment for a short study period so that an absolute TTS reduction magnitude corresponds to a large percentage.

This study tried to shed light on how to transfer the VSL control policies between different scenarios using the transfer learning in the deep reinforcement learning framework and what benefits can be obtained to do so. As the existence of correlations between these scenarios, the transferred DDQN-based VSL policy could work as prior knowledge in the development of optimal VSL policies for new scenarios. As a result, processes of developing optimal VSL policies were greatly shortened, leading to higher time efficiency and data efficiency. Higher time efficiency is especially helpful, when developing a DDQN-based VSL for a large-scale network that requires pretty long training time. Data efficiency is also essential in some practical situations [51]. For example, data of some infrequent scenarios (e.g., adverse weather scenarios) may be deficient, while data of usual scenarios are sufficient. The transfer learning could utilize prior knowledge from data of the usual scenarios, and thereby requires fewer data of infrequent scenarios to obtain adequate training.

Note that the transferability was not compared with that of other VSL strategies (e.g., feedback methods and model predictive control methods) as the transferring mechanisms are different. When transferring the feedback-based VSL, one should conduct traffic flow analysis to decide the target variable value in new scenarios and finely tune the controller gains to maintain control performances, which is time-consuming and require large data support. As for the MPC methods, the traffic flow models should be developed and calibrated for the new scenarios, which requires a lot of workloads. When transferring the DRL-based VSL, we do not need to change the algorithm settings. The DRL can learn the traffic flow implicitly during the training to adapt to different traffic models without extra tuning. In the present study, transfer learning was only used to transfer knowledge between different traffic scenarios. In fact, transfer learning can also be conducted between different but similar traffic control techniques. For example, VSL and ramp metering (RM) could have something in common as they both adjust the flow into a downstream bottleneck to mitigate congestion. Hence, it is highly likely to transfer the pieces of knowledge between VSL and RM. Authors recommend that future research can consider these issues.

## References

[1] M. J. Cassidy and J. Rudjanakanoknad, "Increasing the capacity of an isolated merge by metering its on-ramp," *Transp. Res. B, Methodol.*, vol. 39, no. 10, pp. 896–913, Dec. 2005.

[2] K. Chung, J. Rudjanakanoknad, and M. J. Cassidy, "Relation between traffic density and capacity drop at three freeway bottlenecks," *Transp. Res. B, Methodol.*, vol. 41, no. 1, pp. 82–95, Jan. 2007.

[3] Y. Han, D. Chen, and S. Ahn, "Variable speed limit control at fixed freeway bottlenecks using connected vehicles," *Transp. Res. B, Methodol.*, vol. 98, pp. 113–134, Apr. 2017.

[4] Z. Li, P. Liu, C. Xu, and W. Wang, "Development of analytical procedure for selection of control measures to reduce congestions at various freeway bottlenecks," *J. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 65–85, Jan. 2018.

[5] R. C. Carlson, I. Papamichail, M. Papageorgiou, and A. Messmer, "Optimal motorway traffic flow control involving variable speed limits and ramp metering," *Transp. Sci.*, vol. 44, no. 2, pp. 238–253, May 2010.

[6] R. C. Carlson, I. Papamichail, M. Papageorgiou, and A. Messmer, "Optimal mainstream traffic flow control of large-scale motorway networks," *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 2, pp. 193–212, Apr. 2010.

[7] Y. Han, A. Hegyi, Y. Yuan, S. Hoogendoorn, M. Papageorgiou, and C. Roncoli, "Resolving freeway jam waves by discrete first-order model-based predictive control of variable speed limits," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 405–420, Apr. 2017.

[8] Y. Han, A. Hegyi, Y. Yuan, and S. Hoogendoorn, "Validation of an extended discrete first-order model with variable speed limits," *Transp. Res. C, Emerg. Technol.*, vol. 83, pp. 1–17, Oct. 2017.

[9] H. Liu, L. Zhang, D. Sun, and D. Wang, "Optimize the settings of variable speed limit system to improve the performance of freeway traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3249–3257, Dec. 2015.

[10] R. C. Carlson, I. Papamichail, and M. Papageorgiou, "Local feedback-based mainstream traffic flow control on motorways using variable speed limits," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1261–1276, Dec. 2011.

[11] R. C. Carlson, I. Papamichail, and M. Papageorgiou, "Comparison of local feedback controllers for the mainstream traffic flow on freeways using variable speed limits," *J. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 268–281, Oct. 2013.

[12] G.-R. Iordanidou, C. Roncoli, I. Papamichail, and M. Papageorgiou, "Feedback-based mainstream traffic flow control for multiple bottlenecks on motorways," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 610–621, Apr. 2015.

[13] H.-Y. Jin and W.-L. Jin, "Control of a lane-drop bottleneck through variable speed limits," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 568–584, Sep. 2015.
[14] E. R. Muller, R. C. Carlson, W. Kraus, and M. Papageorgiou, "Microsimulation analysis of practical aspects of traffic control with variable speed limits," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 512–523, Feb. 2015.
[15] G.-R. Iordanidou, I. Papamichail, C. Roncoli, and M. Papageorgiou, "Feedback-based integrated motorway traffic flow control with delay balancing," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2319–2329, Sep. 2017.
[16] Y. Zhang and P. A. Ioannou, "Combined variable speed limit and lane change control for highway traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1812–1823, Jul. 2017.
[17] I. Papamichail, A. Kotsialos, I. Margonis, and M. Papageorgiou, "Coordinated ramp metering for freeway networks—A model-predictive hierarchical control approach," *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 3, pp. 311–331, Jun. 2010.
[18] F. Zhu and S. V. Ukkusuri, "Accounting for dynamic speed limit control in a stochastic traffic environment: A reinforcement learning approach," *Transp. Res. C, Emerg. Technol.*, vol. 41, pp. 30–47, Apr. 2014.
[19] Z. Li, P. Liu, C. Xu, H. Duan, and W. Wang, "Reinforcement learning-based variable speed limit control strategy to reduce traffic congestion at freeway recurrent bottlenecks," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 3204–3217, Nov. 2017.
[20] M. Yang, Z. Li, Z. Ke, and M. Li, "A deep reinforcement learning-based ramp metering control framework for improving traffic operation at freeway weaving sections," in *Proc. 98th Annu. Meeting Transp. Res. Board*, Washington, DC, USA, 2019, pp. 1–8.
[21] C. Lu, J. Huang, L. Deng, and J. Gong, "Coordinated ramp metering with equity consideration using reinforcement learning," *J. Transp. Eng., A, Syst.*, vol. 143, no. 7, Jul. 2017, Art. no. 04017028.
[22] F. Belletti, D. Haziza, G. Gomes, and A. M. Bayen, "Expert level control of ramp metering based on multi-task deep reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 4, pp. 1198–1207, Apr. 2018.
[23] S. Jialin Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[24] J. Ramon, K. Driessens, and T. Croonenborghs, "Transfer learning in reinforcement learning problems through partial policy recycling," in *Proc. 18th Eur. Conf. Mach. Learn. (ECML)*, 2007, pp. 699–707.
[25] M. E. Taylor and P. Stone, "Cross-domain transfer for reinforcement learning," in *Proc. 24th Int. Conf. Mach. Learn. (ICML)*, New York, NY, USA, 2007, pp. 879–886.
[26] M. E. Taylor, S. Whiteson, and P. Stone, "Transfer via inter-task mappings in policy search reinforcement learning," in *Proc. 6th Int. Joint Conf. Auto. Agents Multiagent Syst. (AAMAS)*, May 2007, pp. 156–163.
[27] A. R. Kreidieh, C. Wu, and A. M. Bayen, "Dissipating stop-and-go waves in closed and open networks via deep reinforcement learning," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 1475–1480.
[28] C. F. Daganzo, *Fundamentals of Transportation and Traffic Operations*. Oxford, U.K.: Pergamon, 1997.
[29] Z. Ke, Z. Li, P. Liu, and C. Xu, "A double deep q network-based variable speed limit control to reduce travel time at freeway bottlenecks," in *Proc. 99th Annu. Meeting Transp. Res. Board*, Washington, DC, USA, 2020, pp. 1–20.
[30] M. Papageorgiou and A. Kotsialos, "Freeway ramp metering: An overview," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 4, pp. 271–281, Dec. 2002.
[31] M. J. Cassidy, "Freeway on-ramp metering, delay savings, and diverge bottleneck," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1856, no. 1, pp. 1–5, Jan. 2003.
[32] L. Zhang and D. Levinson, "Ramp metering and freeway bottleneck capacity," *Transp. Res. A, Policy Pract.*, vol. 44, no. 4, pp. 218–235, May 2010.
[33] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transp. Res. B, Methodol.*, vol. 28, no. 4, pp. 269–287, Aug. 1994.
[34] P. Kachroo and S. Sastry, "Traffic assignment using a density-based travel-time function for intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 5, pp. 1438–1447, May 2016.
[35] H. B. Celikoglu, "Dynamic classification of traffic flow patterns simulated by a switching multimode discrete cell transmission model," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 2539–2550, Dec. 2014.
[36] H. B. Celikoglu and M. A. Silgu, "Extension of traffic flow pattern dynamic classification by a macroscopic model using multivariate clustering," *Transp. Sci.*, vol. 50, no. 3, pp. 966–981, Aug. 2016.
[37] R. C. Carlson, I. Papamichail, and M. Papageorgiou, "Integrated feedback ramp metering and mainstream traffic flow control on motorways using variable speed limits," *Transp. Res. C, Emerg. Technol.*, vol. 46, pp. 209–221, Sep. 2014.
[38] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
[39] Z. Li, P. Liu, W. Wang, and C. Xu, "Development of a control strategy of variable speed limits to reduce rear-end collision risks near freeway recurrent bottlenecks," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 866–877, Apr. 2014.
[40] Z. Li, P. Liu, C. Xu, and W. Wang, "Optimal mainline variable speed limit control to improve safety on large-scale freeway segments," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 31, no. 5, pp. 366–380, May 2016.
[41] L. Muñoz, X. Sun, R. Horowitz, and L. Alvarez, "Piecewise-linearized cell transmission model and parameter calibration methodology," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1965, no. 1, pp. 183–191, Jan. 2006.
[42] M. Mauch and M. J. Cassidy, "Freeway traffic oscillations: Observations and predictions," in *Proc. 15th Int. Symp. Transp. Traffic Theory*, 2002, pp. 653–674.
[43] California Department of Transportation. *Caltrans Performance Measurement System (PeMS)*. Accessed: Jul. 15, 2018. [Online]. Available: http://pems.dot.ca.gov/
[44] B. Hellinga and M. Mandelzys, "Impact of driver compliance on the safety and operational impacts of freeway variable speed limit systems," *J. Transp. Eng.*, vol. 137, no. 4, pp. 260–268, Apr. 2011.
[45] A. Ibrahim and F. Hall, "Effect of adverse weather conditions on speed-flow-occupancy relationships," *Transp. Res. Board*, vol. 1457, pp. 184–191, Aug. 1994.
[46] A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2003.
[47] M. Agarwal, T. H. Maze, and R. Souleyrette, "Impacts of weather on urban freeway traffic flow characteristics and facility capacity," in *Proc. Mid-Continent Transp. Res. Symp.*, Ames, IA, USA, Aug. 2005, pp. 1–20.
[48] Y. Kan, Y. Wang, M. Papageorgiou, and I. Papamichail, "Local ramp metering with distant downstream bottlenecks: A comparative study," *Transp. Res. C, Emerg. Technol.*, vol. 62, pp. 149–170, Jan. 2016.
[49] M. Hadiuzzaman and T. Z. Qiu, "Cell transmission model based variable speed limit control for freeways," *Can. J. Civil Eng.*, vol. 40, no. 1, pp. 46–56, Jan. 2013.
[50] M. E. Taylor and P. Stone, "Transfer learning for reinforcement learning domains: A survey," *J. Mach. Learn. Res.*, vol. 10, pp. 1633–1685, Jul. 2009.
[51] A. Lazaric, "Knowledge transfer in reinforcement learning," Ph.D. dissertation, Dept. Electron. Inf., Politecnico di Milano, Milan, Italy, 2008.

**Zemian Ke** received the B.Eng. degree from the School of Transportation, Southeast University, Nanjing, China, where he is currently pursuing the master's degree with the School of Transportation. His research interests include traffic control and traffic flow theory. His team won the First Prize from the Nanjing University Students Transportation Technology Competition in 2016. He won the Excellent Undergraduate Dissertation by the School of Transportation, Southeast University, in 2017.

**Zhibin Li** received the Ph.D. degree from the School of Transportation, Southeast University, China, in 2014. From 2010 to 2012, he was a Visiting Student at the University of California at Berkeley. From 2015 to 2017, he worked as a Post-Doctoral Researcher with the University of Washington and The Hong Kong Polytechnic University. He is currently a Professor with Southeast University. He has authored or coauthored over 60 articles in journals. His research interests include intelligent transportation, traffic safety, data mining, traffic control, and artificial intelligence.

**Zehong Cao** (Member, IEEE) received the B.S. degree in electronics engineering from The Chinese University of Hong Kong, the M.S. degree in electronics engineering from Northeastern University, China, and the Ph.D. degree in information technology from the University of Technology Sydney (UTS). He is a Lecturer (also known as Assistant Professor) with the Discipline of Information and Communication Technology (ICT), University of Tasmania (UTAS), Hobart, Australia, and an Adjunct Fellow of the School of Computer Science, University of Technology Sydney (UTS), Australia. He has published 40+ papers in well-known conferences, such as AAMAS, IJCNN, and the IEEE-FUZZY, and top-tier journals such as the IEEE TRANSACTIONS ON FUZZY SYSTEMS, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, the IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS, the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEM, the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, the IEEE INTERNET OF THINGS JOURNAL, the IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, ACM TOMM, and NeuroImage. He is currently focusing on the capacity of the Human-in-the-Loop machine learning and its applications. His research interests cover brain–computer interface, computational intelligence, and machine learning. He was awarded the UTS Centre for Artificial Intelligence Best Student Paper Award, the UTS Faculty of Engineering and IT Ph.D. Publication Award, and the UTS President Ph.D. Scholarship. He serves as the Leading Guest Editor for the IEEE TRANSACTIONS FUZZY SYSTEMS and the IEEE TRANSACTIONS INDUSTRIAL INFORMATICS, and an Associate Editor for *Neurocomputing*. Since 2019, he has been serving as an Associate Editor for *Scientific Data* and the *Journal of Intelligent Fuzzy Systems*.

**Pan Liu** received the Ph.D. degree in civil engineering from the University of South Florida, Tampa, USA, in 2006. He is currently a Professor with the School of Transportation, Southeast University, Nanjing, China. He has authored or coauthored over 80 articles in transportation journals. His research interests include traffic operations and safety, and intelligent transportation systems. He was a recipient of the Distinguished Young Scientist of NSFC in 2013 and the Award of the Program for New Century Excellent Talents funded by the Ministry of Education, China.